Deliberate practice predicts performance throughout time in adolescent chess players and dropouts: A linear mixed models analysis.

de Bruin, A.B.H.; Smits, N.; Rikers, R.M.J.P.; Schmidt, H.G.

published in
British Journal of Psychology
2008

DOI (link to publisher)
10.1348/000712608X295631

document version
Publisher's PDF, also known as Version of record

British Journal of Psychology (2008), 99, 473–497
© 2008 The British Psychological Society

The British Psychological Society
www.bpsjournals.co.uk

Deliberate practice predicts performance over time in adolescent chess players and drop-outs: A linear mixed models analysis

Anique B. H. de Bruin[1]*, Niels Smits[2], Remy M. J. P. Rikers[1] and Henk G. Schmidt[1]

[1]Erasmus University Rotterdam, Rotterdam, The Netherlands
[2]Free University Amsterdam, Amsterdam, The Netherlands

In this study, the longitudinal relation between deliberate practice and performance in chess was examined using a linear mixed models analysis. The practice activities and performance ratings of young elite chess players, who were either in, or had dropped out of the Dutch national chess training, were analysed since they had started playing chess seriously. The results revealed that deliberate practice (i.e. serious chess study alone and serious chess play) strongly contributed to chess performance. The influence of deliberate practice was not only observable in current performance, but also over chess players' careers. Moreover, although the drop-outs' chess ratings developed more slowly over time, both the persistent and drop-out chess players benefited to the same extent from investments in deliberate practice. Finally, the effect of gender on chess performance proved to be much smaller than the effect of deliberate practice. This study provides longitudinal support for the monotonic benefits assumption of deliberate practice, by showing that over chess players' careers, deliberate practice has a significant effect on performance, and to the same extent for chess players of different ultimate performance levels. The results of this study are not in line with critique raised against the deliberate practice theory that the factors deliberate practice and talent could be confounded.

The necessity of abundant domain-specific experience to acquire exceptional performance has been widely acknowledged in several reviews (e.g. Ericsson, 1996; Ericsson & Lehmann, 1996). Until now, no examples exist of individuals performing at international levels, or improving abruptly without extensive practice in the domain under consideration. In this regard, Simon and Chase (1973) formulated the 10-year rule of expertise, stating that to reach grandmaster level in chess at least 10 years of intense preparation are required. Subsequent studies showed that this rule could

**Correspondence should be addressed to Anique B. H. de Bruin, Department of Psychology, Erasmus University Rotterdam, Rotterdam, 3000 DR, The Netherlands (e-mail: debruin@fsw.eur.nl).*

DOI:10.1348/000712608X295631

be extended to other domains (e.g. mathematics, Gustin, 1985; tennis, Monsaas, 1985; science and arts, Ericsson, Krampe, & Tesch-Römer, 1993). In addition, Ericsson *et al.* (1993) articulated that especially practice of high quality is beneficial to performance improvement. They argue that 'deliberate practice' is the key activity that determines progress. Deliberate practice is defined as practice that (1) is primarily directed at performance improvement, (2) is of adequate difficulty, (3) involves informative feedback, and (4) provides ample opportunity for repetition and correction of errors. As such, deliberate practice can be opposed to usual work activities, which do not possess the same degree of control and are generally more directed at problem solving than at performance improvement. Deliberate practice requires full concentration and, therefore, can only be performed for a few hours per day.

To test their theory, Ericsson and colleagues (1993) asked student violinists of several performance levels to estimate how many hours per week they had practiced alone playing the violin for each year since starting. The retrospective practice estimates showed a steady increase in practice since starting age until the time of the study. Moreover, the best students had averaged 7,401 hours of practice alone, whereas good students had practiced significantly less: 5,301 hours. The cumulative number of hours of practice alone was lowest for the participants studying to become music teachers: 3,420 hours. Ericsson and colleagues concluded that practice alone, considered as a primary form of deliberate practice in the music domain, strongly contributed to expertise in music.

Since their initial paper (Ericsson *et al.*, 1993), the positive relation between accumulated amount of deliberate practice and performance has been shown in diverse fields as soccer (Helsen, Starkes, & Hodges, 1998), martial arts (Hodge & Deakin, 1998), triathletes and swimmers (Hodges, Kerr, Starkes, Weir, & Nananidou, 2004), chess (Charness, Krampe, & Mayr, 1996; Charness, Tuffiash, Krampe, Reingold, & Vasyukova, 2005), music (Sloboda, Davidson, Howe, & Moore, 1996), and teaching (Dunn & Shriner, 1999). In these studies, current expertise level was always monotonically related to accumulated amount of deliberate practice. In contrast, practice measures that were not considered as deliberate practice, as tournament play or practising for fun, did not contribute to current expertise level. For example, Charness *et al.* (1996) analysed in chess to what extent several practice data and biographic variables, as starting and club joining age, predicted current chess rating. Their results showed that serious chess study alone, but not analysing chess games with others or playing in chess tournaments, contributed significantly to current chess rating. Moreover, starting and club joining age were not significant predictors of current chess rating, as was also the case for having a coach. In a subsequent study, Charness and colleagues (2005) again found that cumulative hours of serious study alone was the strongest predictor of current chess rating. Moreover, total years of private instruction and current hours of serious chess study alone also contributed to performance. However, a possible drawback of their study is that Charness and colleagues failed to evaluate the reliability of the retrospective estimates of practice hours, by comparing them, for example, to self-report diaries (e.g. Ericsson *et al.*, 1993; Sloboda *et al.*, 1996; Starkes, Deakin, Allard, Hodges, & Hayes, 1996). Although many studies have replicated the basic finding of a monotonic relation between accumulated deliberate practice and current performance, few have addressed critiques or questions raised over the years regarding some of the assumptions of the deliberate practice theory.

Monotonic benefits assumption

As noted by Hodges and colleagues (2004), deliberate practice research has mainly focused at predicting differences in performance *between* skill levels, whereas a stronger test would be to examine whether differences in amount of deliberate practice *within* skill levels show the same association with performance. Moreover, research on deliberate practice and expertise development has been directed at explaining and predicting current skill level, such as swimming times (Hodges *et al.*, 2004) or ELO ratings in chess (Charness *et al.*, 2005), whereas little research has taken into account the development of performance, and concurrent engagement in deliberate practice, over time. Research is needed that studies the relation between deliberate practice and performance longitudinally to enable analysis of certain assumptions that can extend the scope of the theory (Hodges *et al.*, 2004). For example, the deliberate practice theory was initially put forward as a theory explaining performance of top-level experts of national or international level. However, the basic assumption by Ericsson and colleagues (Ericsson & Charness, 1994; Ericsson *et al.*, 1993), the 'monotonic benefits assumption', entails that accumulated deliberate practice is monotonically related to current performance. That is, there is a direct, monotonically increasing relationship between amount of deliberate practice and performance over time, which will approach an asymptote as individuals reach their performance optimum. This would imply that within individuals, at any point throughout their career, amount of deliberate practice is associated with performance. The theory does not assume that practice can have delayed effects on performance that would predict zero-order associations between practice and performance at specific time measurements throughout career. However, this part of the assumption has never been studied longitudinally. Therefore, research is required that studies the relation between deliberate practice and performance, not only for current skill level, but also over time.

Given the limited access to longitudinally assessed performance measures, studies addressing the tenability of the monotonic benefits assumption over experts' careers are rare. This assumption has mainly been tested by comparing accumulated practice hours between individuals of varying skill levels (e.g. Baker, Côté, & Deakin, 2005; Helsen *et al.*, 1998; Hodge & Deakin, 1998; Sloboda *et al.*, 1996). However, because these studies did not include performance measures that were longitudinally assessed, they were unable to draw any conclusions about the influence of deliberate practice on performance at earlier career points. Since the monotonic benefits assumption has both theoretical and practical implications that could affect training of non-experts, research is needed that assesses its validity when taking into account the longitudinal development of performance.

A question related to the longitudinal aspect of the monotonic benefits assumption is at what point in a career practice is most determining of performance. The monotonic benefits assumption predicts a direct relation between practice and performance at all career stages. However, until now, research concerning this issue has shown equivocal results. Originally, Krampe and Ericsson (1996) showed that deliberate practice over the last decade was most strongly related to current performance in music. Moreover, according to Baker, Côté, and Abernethy (2003), practice does not have to be sport specific until later in the careers of athletes to exert an effect on performance. That is, early specialization is not necessary and diversity in training may be more beneficial, because it promotes the development of more flexible perceptual-motor skills. However, these data are not uncontroversial. Hodges *et al.*'s (2004) analysis in swimming and running failed to replicate these findings: an analysis that included the most recent practice years as a predictor of performance explained less variance than a model including recent as well as more distant practice years. Moreover, Ward and colleagues (Ward, Hodges, Williams, & Starkes, 2004) showed that sport-specific team practice was the strongest contributor to soccer expertise from an early age on. Investment in non-sport specific play-like activities at an early age did not differ between elite and subelite players. However, the effect of career stage on the relation between practice and performance has never been studied in chess. Because chess, in contrast to physical sports, does not rely on general motor skills but mainly on specific cognitive skills, the effect of early sport-specific practice might be less prominent in that domain. Analysing the relation between deliberate practice and performance over time in chess could provide further insight into this issue.

Gender differences in cognitive domains

A theme that has emerged more recently in research related to deliberate practice, is the effect of gender when explaining variance in performance. Hodges and colleagues (2004) showed that gender, after controlling for differences in deliberate practice, is predominantly a determinant in sports that are primarily of anaerobic nature (e.g. depending on physical factors as muscle strength and distribution of fat in the body), and less so in aerobic activities as distance running or swimming. However, in the case of cognitive tasks, where muscle strength and fat distribution do not exert an influence, the deliberate practice account does not recognize any genetically based differences between males and females, and asserts that amount of deliberate practice is the foremost determinant of performance.

Although a cognitive task, performance differences between males and females have been widely acknowledged in chess (Howard, 2004). Explanations that have been proposed for this discrepancy, concern genetic variation between males and females that could cause specific ability differences (i.e. visuospatial skills) or social factors that lead to different socialization environments for boys and girls. While research has shown that males on average tend to score higher than females on tests of visuospatial ability (e.g. Benbow, Lubinski, Shea, & Eftekhari-Sanjani, 2000), there is little evidence for an association between visuospatial skill and chess performance (Doll & Mayr, 1987; Waters, Gobet, & Leyden, 2002). Frydman and Lynn (1992) showed that young chess players had higher performance IQ scores than non-chess players, but could not prove that this was due to higher scores on measures of visuospatial ability. Therefore, no clear evidence exists for a relation between chess skill and performance on visuospatial tasks, which, in-turn, could account for gender differences in chess skill.

The explanation for gender differences in chess focusing on social causes describes how factors such as less female role models and fewer opportunities to become high achievers cause lower female performance (Howard, 2004). Assuming that in the last decades these social factors have diminished considerably, Howard (2004) sets out to explore whether this development is also reflected in a convergence of performance between males and females in chess. He analysed the FIDE (World Chess Federation) ratings of all males and females since the start of publication in 1970. His data suggest little evidence for convergence of performance in chess between males and females. For example, when looking at the top ranked players, males still score more than a standard deviation above females.

There are, however, a number of factors that allow for an alternative explanation of Howard's results. First, Howard (2004) does not take into account the different participation rates of males and females in chess (Bilalić & McLeod, 2006; Chabris & Glickman, 2006; Charness & Gerschak, 1996). That is, the mere fact that more males than females compete in chess increases the statistical probability that males will become top-rated players. When participation of males and females is equal, both sexes should have the same opportunities to reach the top. This hypothesis was recently confirmed by research of Chabris and Glickman (2006). They showed that, when selecting pairs of males and females matched on chess rating, age and number of games played, these groups showed similar development in chess ratings and comparable drop-out rates in the following 10 years. However, the pairs of matched males and females were formed *post hoc*. It is possible that matching them on age, rating, and number of games played also led to matching on other variables as learning rate and intelligence, which might have confounded their conclusions. To further test the tenability of the participation rate hypothesis as an explanation for male–female differences in chess, it would be relevant to study it in an unmatched sample of chess players where the proportion of males is naturally similar to the proportion of females.

A second alternative explanation for Howard's (2004) data concerns the amount of practice males and females have engaged in. For the group of players that first obtained a FIDE ranking between 1985 and 1989, the males had played 57% more chess games than the females (for the top 100 players, this was even 300%). Moreover, only 39% of the females had an active chess past after 1999, as opposed to 67% of the males, which indicates that the males, for unknown reasons, stayed longer on the list, and probably were more active in chess. Therefore, an alternative explanation not tested by Howard concerns a difference in practice hours between the sexes. It is possible that males dedicate more time to deliberate practice than females, which could contribute to differences in performance. This is substantiated by research from Baker *et al.* (2003), who demonstrated that to become an expert in either basketball, netball, or field hockey, females put in less than half the number of deliberate practice hours that males put in (2,543 hours vs. 5,159 hours, respectively). Apparently, to compete at an international level in these sports requires less time investment from females than from males. Note, however, that research in darts has shown that amount of deliberate practice does not explain male–female performance differences (Duffy, Baluch, & Ericsson, 2004). It seems that in darts other factors besides deliberate practice contribute to performance. To examine this relation in chess, research is needed that simultaneously takes into account gender and deliberate practice to explain performance.

A further shortcoming in Howard's study is that he combines information from all players that first appeared on the rating list between 1985 and 1989, and he does not control for when people entered or disappeared from the list. In his approach, different background, different entry rating, and a number of uncontrolled social factors might account for part of the variance in ratings. A stronger evaluation of gender differences in career development would be to select a group of chess players of about equal playing strength with similar training opportunities, and to examine whether males and females within that group develop distinctively.

Explaining drop-out in promising chess players

A fundamental critique raised against the deliberate practice theory is that the factors talent and deliberate practice might be confounded (Sternberg, 1996). That is, because research on deliberate practice has focused on experts who have reached the top, the deliberate practice investments of those who have dropped out along the way have been disregarded. Ignoring those who have not made it, and focusing on those who have, might lead to an overestimation of the relation between deliberate practice and expertise. For instance, if those who dropped out have put in a comparable number of hours of deliberate practice as those who reached the top, this would contradict the suggested monotonic relation between deliberate practice and performance. Perhaps talented individuals use deliberate practice more readily and thus proceed faster on the road to expertise (Sternberg, 1996). To test this hypothesis, research should examine how equal investments in deliberate practice affect performance improvement of drop-outs and persisters in chess. If such an analysis reveals that in both groups the monotonic benefits assumption holds, this proves that the often-observed relation between deliberate practice and achievement is not an artefact of subject characteristics (e.g. a predisposition to readily perform deliberate practice). By contrast, this would suggest that regardless of performance level (i.e. persisting or dropping out) equal amounts of deliberate practice improve performance to equal extents.

In sum, although the positive relation between accumulated deliberate practice and performance in high-level experts has been documented in diverse fields, certain issues need to be addressed to examine the scope of the deliberate practice theory. For example, it is at present unclear to what extent the monotonic benefits assumption holds at all career stages. If deliberate practice proves beneficial at early stages in career development, this will likely influence non-expert training and teaching. Second, given the lack of studies relating deliberate practice to gender and performance differences in chess, research is needed that examines these factors simultaneously. Finally, the field is in need of research that analyses the extent to which the effect of deliberate practice is comparable across groups of varying performance success, and not dependent on a predisposition to effectively engage in deliberate practice. This can be achieved by taking into account persisters and drop-outs in chess.

The purpose of the present study was to test these three assumptions. To this end, the relation between (a) deliberate practice (i.e. serious chess study alone and serious chess play, Charness *et al.*, 1996), (b) gender, and (c) the development of chess performance (measured through chess ratings) was examined over time, by using a linear mixed models analysis. While Hodges *et al.*'s (2004) study provides evidence that within skill level deliberate practice hours are related to current performance, to our knowledge, no research has studied this relation over time. To test these relations, we asked a group of elite, adolescent chess players, who participated in the highest level of national chess training of the Dutch Chess Federation, and players who had participated but dropped out, to complete a questionnaire regarding their chess practice activities since the time they started to play chess seriously. These groups were both qualified for the highest level of chess training provided in The Netherlands. The availability of the chess ratings provided several times a year by the Dutch Chess Federation enabled analysis of the longitudinal relation between chess ratings, deliberate practice, gender, and the decision to quit or to proceed seriously playing chess. Because both the deliberate practice estimates and the chess ratings were time dependent, we were able to analyse per time measurement to what extent deliberate practice and performance were related. In contrast to previous deliberate practice studies in chess (e.g. Charness *et al.*, 1996, 2005), we assessed the reliability of the deliberate practice estimates by asking a subset of the participants to complete a self-report diary for 3 consecutive weeks.

The monotonic benefits assumption predicts an approximately linear relation between deliberate practice and performance for all career stages, except possibly for the final stage in which a performance asymptote is reached. Given the relatively young age of our participants, we predicted that deliberate practice hours would linearly contribute to their chess ratings, irrespective of point in career or time passed since starting to play chess. Furthermore, we predicted that those who were still in the national training would gain higher chess ratings over time, given the higher number of deliberate practice hours they put in, compared to those who dropped out. We predicted that a different line of reasoning would apply when taking into account the factor gender. Since no evidence exists that playing chess relies on capacities associated with gender specific advantages, and given that previous research suggests that males and females tend to differ in how much time they dedicate to practice, we predicted that gender would not contribute to performance, once deliberate practice was controlled for.

Method

Sample

A total of 81 adolescent chess players (30 girls, 51 boys) participated in this study (*M* age = 16.19, *SD* = 2.75, range 12–23 years). Note that in this sample the proportion of males and females was naturally comparable, in contrast to the artificially matched sample in Chabris and Glickman (2006). These chess players were either in the 2003–2004 selection of the national chess training of the Dutch Chess Federation, or were selected in earlier years, but had previously decided to quit (*M* chess rating at time of test = 1, 944, *SD* = 259). The average Dutch chess rating of a competitive adult chess player is around 1,700. An individual with a chess rating higher than 2,000 is considered an expert player. This holds for less than 10% of those who receive a rating. Every year, the Dutch Chess Federation selects about 10 adolescents from the top-performing chess players in The Netherlands for national training, based on chess ratings and information from regional coaches. To ensure that memories of the training years for those who had quit were fairly recent, we only selected the 24 players who had quit between 1999 and 2004. Two of these were not willing to take part. For the remaining group, 18.2% dropped out in 2003, 36.4% in 2002, 13.6% in 2001, 18.2% in 2000, and 9.1% in 1999. Of the 59 participants tested in June 2004 (92.2% of the total national training group at that time), 11 players later decided not to return to the national training after the summer. Therefore, these individuals were in the analyses taken into the group that had dropped out. All in all, the group that still received national training (hereafter referred to as 'persisters') consisted of 48 participants (*M* age = 15.13 years, *SD* = 2.14), whereas the group that had quit the national training during or before the summer of 2004 (hereafter referred to as 'drop-outs') consisted of 33 participants (*M* age = 17.77 years, *SD* = 2.85). These drop-outs were never asked to leave, but, when invited for the first national training after the summer, declined to return. The age at which the drop-outs had quit was on average 16.12 years (*SD* = 2.02). This sample is highly representative of the national training group, given that 92.2% of the national training participated and 91.7% of the drop-outs took part. All participants received a small financial compensation after completion of the study.

Materials

Participants completed a paper-and-pencil questionnaire that consisted of four sections. Data of the third and fourth part were collected for a different study on the relation between achievement motivation and chess performance (De Bruin, Rikers, & Schmidt, 2007). The first part consisted of biographic information questions (age when learning to play chess, number of chess books, etc.). The second part inquired about players' cumulative hours of (1) serious chess play against other chess players and (2) serious analysis of chess games alone (serious chess study alone, cf. Charness *et al.*, 1996, 2005). Chess games over the internet were not taken into account in this measure of serious chess play, because, for this specific sample, these are considered playful interaction.

Apart from the questionnaire, participants' chess ratings were collected with help of the Dutch Chess Federation, from the start of their chess career. These ratings are calculated in the same manner as the ratings by the World Chess Federation. The only difference between Dutch ratings and FIDE ratings is that the former are also based on regional Dutch tournaments, whereas the latter are solely based on official FIDE tournaments. The Dutch ratings, therefore, tend to be somewhat lower than the FIDE ratings. However, since we were not interested in the absolute level of the ratings, but in their development over time, we did not transform the Dutch ratings to FIDE ratings. Until 2000, the Dutch Chess Federation published chess ratings twice a year. In 2001, ratings came out three times, and as of 2002 these ratings were published four times a year. Depending on when the participant started, two, three, or four measurements are available per year.

Procedure

The persisters filled out the questionnaire individually during a national chess-training weekend in June 2004. Since the drop-outs no longer attended the national trainings, a research assistant visited them at home and asked them to complete the questionnaire.

Analysis

The cumulative number of hours spent on serious chess study alone and on serious chess play against others was calculated by multiplying the weekly estimates by 52 and summing the total hours for each year. This probably overestimates the total cumulative hours of chess study alone and serious chess play somewhat, but provides a consistent pattern across individual chess players.

Reliability of retrospective estimates

To maximize reliability of the retrospective estimates, we took a number of precautions:

(1) We ensured that participants' memories of their practice activities were as recent as possible. Therefore, chess players who had dropped out of the national training before 1999 were not allowed to participate. To assess possible within-group differences in memory for practice activities, we analysed whether differences existed in reported number of hours of deliberate practice (sum of total hours of serious chess play, and serious chess study alone) between those who had stopped longer ago (i.e. before the year 2002) to those who stopped in the more recent past (i.e. during, or after the year 2002), using a oneway ANOVA. For both serious chess play and serious chess study alone, no differences between these groups were found, both Fs < 1.

(2) Following the encoding specificity principle formulated by Tulving and Thomson (1973), we encouraged participants at the time of retrieving memories of practice to reinstate the mental context at the time of encoding. This technique is also

applied in the cognitive interview (Geiselman *et al.*, 1984) to increase the accuracy of eyewitness memories, and is by some considered the most effective component of the cognitive interview (Memon & Bull, 1991). We asked participants before filling in the number of hours they had spent at a particular age on serious chess play and serious chess study alone, to first attempt to covertly retrieve the following details for themselves, which could act as retrieval cues that would enhance memory for chess playing at that age:

- What grade were you in? Who was your teacher at that age?
- Where were you living at that time? What did your room look like?
- What other hobbies did you have besides chess?
- How was your leisure time distributed over the hobbies you had at that time?
- How did chess fit into your life at that time? Were you in a chess club? How much time did you spend on chess? Were you playing any tournaments?

Since these questions were intended to stimulate participants' retrieval of relevant memories, it was most natural for them to do so covertly. Therefore, we have no data to present on their answers to these questions.
A subsample of the original participants (36 persons: 20 persisters, 16 drop-outs, 44.4% of original sample) completed an internet chess diary for 3 consecutive weeks, which allowed us to calculate the correlation between retrospective deliberate practice estimates of the last year and actual practice hours as indicated by the diary. Although this technique is commonly applied in deliberate practice research, it has not been used in deliberate practice research in chess (Charness *et al.*, 1996, 2005; Sloboda *et al.*, 1996).

Linear mixed models analysis
In longitudinal data, correlations are typically observed between dependent measurements. To handle this type of data, regular statistical analyses, such as the summary statistic method, do not apply [see Omar, Wright, Turner, and Thompson (1999), for an evaluation of techniques to analyse repeated measures]. Moreover, regular repeated measures analyses of variance typically cannot cope with missing data, and only take into account participants with complete data, who might not be representative of the full data set. Also, these analyses estimate group effects, and provide no insight into how individuals develop over time. For these reasons, the mixed-effects regression models have become increasingly popular to model longitudinal data. These models include random regression effects that account for the influence of participants on repeated measurements, and thereby enable analysis of individual development over time. Also, in this type of regression analyses, individuals with incomplete data (i.e. due to drop-out) can be included in the analysis, on the assumption that these data are missing at random.

In general, these mixed-effects regression models consist of a within-subjects (level 1) model, and a between-subjects (level 2) model (Laird & Ware, 1982). Consider the following equation that describes the within-subjects model (Greek letters denote population parameters, whereas Arabic letters refer to parameters of an individual):

$$y_{ij} = b_{0i} + b_{1i}t_{ij} + \varepsilon_{ij}$$

This formula describes measurement y of individual i ($i = 1, 2, \ldots N$ participants) on occasion j ($j = 1, 2, \ldots N$ occasions). This model represents the influence of the independent variable time (denoted t) on the outcome variable y for individual i.

This outcome is further affected by his/her initial level b_{0i} and an error term ε_{ij}. The level 2 (between-subjects model) would look like

$$b_{0i} = \beta_0 + v_{0i}$$

$$b_{1i} = \beta_1 + v_{1i}$$

This model indicates that the individual i's outcome is determined by the population initial level β_0 plus a unique contribution for individual i, namely v_{0i}. In addition, in the general model, the individual growth is determined by both the population increase (β_1) plus the unique contribution for individual i, (v_{1i}). Combining the level 1 and 2 models, the complete model with random intercept and random slope is formulated as follows:

$$y_{ij} = \beta_0 + v_{0i} + \beta_{1i}t_{ij} + v_{1i}t_{ij} + \varepsilon_{ij}$$

This model assumes that the variances of the errors are normally distributed. Visual, exploratory inspection of our data revealed that individuals mainly varied in start chess rating (i.e. the intercept) and less in growth rate. For instance, for all participants, chess ratings increased at a similar rate over time. Therefore, we decided to apply a fixed slope and only estimate a random intercept. Consequently, in the models that are used in the present study, the unique contribution v_{1i} will not be modelled, that is the individual slope (b_{1i}) is equal to the population slope (β_1). Moreover, the participants in this study were too young to have reached the asymptote of their optimal performance level (Ericsson & Charness, 1994; Ericsson *et al.*, 1993): chess players obtain their highest ratings usually not until their 30 seconds or even 40 seconds (Charness & Bosman, 1990; Roring & Charness, 2007). Therefore, we do not only expect a monotonic relation between practice and performance within individuals' careers, but we also assume that this relation can be approached by a linear curve for this age group. That is, in the tested models the monotonic benefits assumption was translated as a linear relation between practice and subject variables on one hand and chess ratings on the other hand.

To examine differences between persisters and drop-outs in growth rate, we incorporated the factor persistence (1 = persisters, 0 = drop – outs) as an independent variable in the analysis. We tested several models to examine which variables should be included to explain the data best. To assess whether a predictor that was added to the model increased the explained variance, we used the Akaike information criterion (AIC; Akaike, 1973). The AIC determines the maximum likelihood of a candidate model, but adjusts it by the number of parameters that are estimated (Burnham & Anderson, 2002). As an additional criterion, we also report the Bayesian information criterion (BIC; Burnham & Anderson, 2002), as this fit measure is insensitive to sampling size, unlike the AIC. The model with the lowest AIC and BIC represents the best model. Because the raw AIC and BIC values are difficult to interpret in terms of the probabilities of the tested models, we transformed these measures to Akaike weights and Schwartz weights, respectively. This enabled us to estimate the relative probability of each model within the set of tested models. That is, the sum of these probabilities amounts to 1 [for a detailed explanation of the calculation of these weights, see Wagenmakers and Farrell (2004)]. Moreover, for nested models, we performed the log likelihood ratio test, and compared the difference in log likelihood between the model with and the model without the added predictor to a Chi square distribution, with the number of added parameters as the degrees of freedom. Finally, we report the significance of the p-values of the coefficients in the model.

Model A1: Deliberate practice and performance over time

First, we tested a model that included time, and the two time-varying covariates serious chess play, and serious chess study alone as predictors of chess rating.

$$y_{ij} = \beta_0 + v_{0i} + \beta_1 T_{1ij} + \beta_2 P_{2ij} + \beta_3 S_{3ij}$$

In this formula, β_0 and β_1 denote the population intercept and slope, whereas v_{0i} indicates the individual variation from the population intercept. The parameter y_{ij} refers to the outcome on measurement y (chess rating) for individual i on occasion j. Variable T_1 refers to time, P_2 refers to serious chess play, and S_3 to serious chess study alone. Time was coded in monthly intervals, starting when players received their first official chess rating. Since chess players were asked to estimate their practice activities *per year*, whereas chess ratings were available two, three, or four times a year, different chess ratings within a year were coupled with one value of serious chess study alone and one value of serious chess play. We preferred this method above asking chess players to estimate practice hours in 3- or 4-month intervals, as this would probably lead to a decrease in accuracy of estimations. For 78 chess players we had information about chess ratings, serious chess study alone, and serious chess play in overlapping time intervals. Therefore, all the tested models were run on 78 participants.

Model A1 implies a linear relation between amount of deliberate practice and chess ratings. To examine the monotonic benefits assumption more, we also ran a model (Model A2) in which the interaction between time and either serious chess play, and time and serious chess study were entered. This allowed us to assess to what extent the relation between deliberate practice and chess ratings changed over time. It could be that the relation between deliberate practice and performance is not linear, but changes after a certain amount of time (e.g. practice early on has a smaller effect on performance than recent practice).

Model B: Including persistence and the interaction between persistence and time

To examine whether the factor persistence (i.e. comparing those who eventually dropped out of the national training, and those who persisted) affected chess ratings, we included the variable persistence (persisters = 1, drop–outs = 0) in the analysis. This variable indicates to what extent differences in chess ratings are due to pre-existing differences between those who ultimately drop-out, and those who persist. Also, to examine whether the influence of the factor persistence changes over time, we added the interaction between persistence and time as a predictor. The model was formulated as follows:

$$y_{ij} = \beta_0 + v_{0i} + \beta_1 T_{1ij} + \beta_2 P_{2ij} + \beta_3 S_{3ij} + \beta_4 D_{4ij} + \beta_5 D_{4ij} T_{1ij}$$

In this model, D indicates whether participants persisted or dropped out of the national training. To assess whether drop-outs and persisters benefited differentially from deliberate practice, we ran two extra models that incorporated either the interaction between persistence and serious chess play (Model B2), or persistence and serious chess study alone (Model B3).

Model C: Effect of gender
The last factor we were interested in was to what extent males and females differ in chess rating development, after the factor deliberate practice is accounted for. To this end, we entered gender, and the interaction between gender and time in the analysis. The former was entered to account for differences in start chess rating. The latter was added to assess whether males' and females' chess ratings developed differently over time. The tested model can be summarized as follows:

$$y_{ij} = \beta_0 + v_{0i} + \beta_1 T_{1ij} + \beta_2 P_{2ij} + \beta_3 S_{3ij} + \beta_4 D_{4ij} + \beta_5 D_{4ij} T_{1ij} + \beta_8 G_{5ij} + \beta_9 G_{5ij} T_{1ij}$$

In this model, G (0 for females, 1 for males) stands for gender. We also ran separate analyses to study the possible differential effect of deliberate practice on gender, by either incorporating the interaction between gender and serious chess play (Model C2), or gender and serious chess study alone (Model C3). To provide the results for these models, we used version 3.0 of the NLME library (an acronym for non-linear mixed effects) for the statistics package R, described in detail by Pinheiro and Bates (2000).

Results

Biographic information

Means and standard deviations for biographic information are represented in Table 1. Confidence intervals for differences between persisters and drop-outs are also reported in Table 1. While most of the biographic information did not differ between persisters and drop-outs (e.g. starting age playing chess, starting age formal training), the total number of chess trainers had was higher for the persisters than for the drop-outs. Persisters had had on average 2.01 more chess trainers than the drop-outs. Persisters owned on average 27.17 more books and 1.68 more CDs than drop-outs. The variables that differed between males and females were more numerous (see Table 2): Males owned on average 44.42 more chess books, 3.12 more chess CDs, and had had 2.06 more chess trainers. Moreover, males had obtained 959.08 more hours of serious chess play, and 1,292.11 more hours of serious chess study. Finally, females received their first chess rating 1.00 year later than males.

Preliminary analyses on chess ratings and deliberate practice

The correlation between accumulated hours of serious chess study alone and most recent chess rating was $r = .45$ ($N = 73$, 95% CI = .24–.62), whereas the correlation between accumulated hours of serious chess play against others and most recent chess rating was $r = .42$ ($N = 75$, 95% CI = .21–.59). In addition, the mean chess rating of the persisters was higher than the drop-outs' chess rating. Mean chess ratings of the persisters was 1,986 ($SD = 266$), mean chess rating of the drop-outs was 1,868 ($SD = 212$, 95% CI of the difference = 26.48–221.81). As to the effect of gender, males had a significantly higher chess rating than females. Males had a mean chess rating of 1,864 ($SD = 140$), whereas females had a mean chess rating of 1,661 ($SD = 140$, 95% CI of the difference = 139.24–290.04).

Reliability of retrospective estimates

We calculated the mean weekly hours of serious chess study alone and serious chess play based on 3 consecutive weeks of diary reports. These means were correlated, using

Table 1. Means and standard deviations of biographic variables for persisters and drop-outs, and 95% confidence interval of difference between persisters and drop-outs

	Persisters ($N = 48$)		Drop-outs ($N = 33$)		
Question	Mean	*SD*	Mean	*SD*	95% confidence interval
Age starting to play chess	6.31	1.70	5.80	1.31	−0.18–1.22
Age first serious about chess	8.65	2.58	7.94	1.89	−0.34–1.75
Age first formal chess instruction	8.10	2.17	8.15	2.23	−1.05–0.94
Age first chess rating	11.83	1.29	12.09	1.51	−0.88–0.37
Mean chess rating	1,986	266	1,868	212	26.48–221.81
Total number of chess trainers	9.25	5.93	7.24	3.24	−0.25–4.26
Number of chess books	58.17	85.17	31.00	36.27	−5.61–59.95
Number of chess CDs	5.88	4.76	4.20	4.45	−0.43–3.80
Number of chess tournaments played per year[a]	21.01	13.64	21.65	13.46	−6.83–5.55
Total number of hours serious chess play	2,309.15	1,782.19	2,263.95	1,608.59	−729.37–819.76
Total number of hours serious chess study alone	1,848.45	2,415.68	1,452.31	1,518.35	−575.10–1,367.38

[a] It is evident that the persisters nowadays play more chess tournaments than the drop-outs. Therefore, we asked the drop-outs how many tournaments they played when they still received national training, to be able to compare chess activities during the national training period. That number is provided in the current Table.

Table 2. Means and standard deviations of biographic variables for males and females, and 95% confidence interval of the difference between males and females

	Males ($N = 51$)		Females ($N = 30$)		
Question	Mean	*SD*	Mean	*SD*	95% confidence interval
Age starting to play chess	6.14	1.65	6.03	1.41	−0.84–0.61
Age first serious about chess	8.48	2.35	8.15	2.33	−1.40–0.74
Age first formal chess instruction	8.48	2.12	7.53	2.18	−1.94–0.04
Age first chess rating	11.57	1.46	12.57	0.97	0.40–1.59
Mean chess rating	1,864	140	1,661	140	139.24–290.04
Total number of chess trainers	9.20	5.85	7.13	3.06	−0.23–4.36
Number of chess books	64.31	84.68	19.90	21.29	12.42–76.41
Number of chess CDs	6.39	5.43	3.267	1.98	1.06–5.18
Number of chess tournaments played per year	19.66	13.26	24.18	13.63	−1.78–10.81
Total number of hours serious chess play	2,638.17	1,901.25	1,679.09	1,055.63	195.84–1,722.32
Total number of hours serious chess study alone	2,154.85	2,477.35	862.74	603.51	341.80–2,242.42

a Pearson product-moment correlation, with the retrospective weekly estimate for these two variables for the current year of practice in the questionnaire. For serious chess play, this correlation was .74 ($N = 36$, 95% CI = .54 - .86), whereas, for serious chess study alone, this was .60 ($N = 36$, 95% CI = .34 - .78). These correlations are comparable to those found in previous studies (e.g. Hodges & Starkes, 1996; Hodges *et al.*, 2004). For serious chess play, mean weekly retrospective estimate was 5.36 hours ($SD = 5.05$), whereas the diary mean for this variable was 6.31 hours ($SD = 6.10$). For serious chess study alone, mean weekly retrospective estimate was 4.38 ($SD = 6.23$), and diary mean was 3.70 ($SD = 3.79$).

Mixed-effects regression models
Since we entered several variables into the analyses that showed relatively high intercorrelations, collinearity among variables could have threatened our results. However, a simulation study by Shie and Fouladi (2003) showed that collinearity did not effect estimation of coefficients of linear mixed models. Therefore, it is unlikely that intercorrelation among predictors biased our estimates. In Table 3, these intercorrelations and their 95% confidence intervals are presented. Moreover, the analyses reported below were also run by centering the predictors in the models. As centering led to equal conclusions in all of the models, we report the non-centered predictors in all analyses.

Model A: Deliberate practice and performance over time
Model A1 examined to what extent deliberate practice contributed to chess ratings, when taking into account the time dependency between these variables. In Table 4, the results for the corresponding regression weights (β_0–β_3) are provided, together with the variance of the random intercept, and the log likelihood of the presented models. This Table shows that all regression weights were significant. Model A2 tested whether

Table 3. Correlations among variables entered in the linear mixed models analyses (95% confidence interval between brackets)

Variables	Age	Sex	Persistence	Serious study	Serious play	Chess rating
Age						
Sex	−.02 (−.08 to .04)					
Persistence	−.42 (−.47 to .37)	−.03 (−.09 to .04)				
Serious study	.04 (−.03 to .10)	.31 (.25 to .36)	.18 (.12 to .24)			
Serious play	.04 (−.02 to .10)	.21 (.15 to .27)	.17 (.11 to .23)	.52 (.47 to .56)		
Chess rating	.39 (.33 to .44)	.37 (.32 to .43)	.03 (−.03 to .10)	.41 (.35 to .46)	.38 (.32 to .43)	
Time	.44 (.39 to .49)	.13 (.07 to .19)	−.10 (−.17 to −.04)	.25 (.19 to .31)	.14 (.08 to .20)	.65 (.62 to .69)

Table 4. Estimated coefficients of model A (and standard errors), which tests the influence of serious chess play and chess study on chess performance over time

	Model A1		Model A2	
Coefficient	Estimate	*SE*	Estimate	*SE*
Fixed parameters				
β_0 (constant)	1,575.18**	20.33	1,561.72**	22.29
β_1 (T_1, time)	6.78**	0.12	7.01**	0.20
β_2 (P_2, serious play)	0.06**	0.02	0.06*	0.03
β_3 (S_3, serious study)	0.12**	0.02	0.17**	0.03
$B_4(P_2 \times T_1)$			0.00	0.00
$B_5(S_3 \times T_1)$			−0.01*	0.00
R	0.710		0.710	
Random parameters				
σ^2_{vi0}	166.06	65.85	168.40	65.92
Number of parameters	5		7	
AIC	11,207.05		11,234.11	
BIC	11,236.27		11,237.07	
Log likelihood	−5,594		−5,592	

Note. T is the number of months from study entry at which the measurements were taken. P_2 (serious play) and S_3 (serious study) were estimated per year. Number of participants = 78; number of measurements = 968. *$p < .05$; **$p < .01$. The log likelihood estimates are based on maximum likelihood estimations, which only included the fixed parameters. Therefore, these differ to a small extent from what would be expected based on the AIC and BIC values. BIC was based on the total number of measurements over all participants (968).

adding the interaction between serious chess play and time, and between serious chess study and time provided a better explanation for the data than the Model A1 that included only linear effects. Although the interaction between serious chess study and time was significant, the AIC and BIC of Model A2 were higher. A likelihood ratio test with two degrees of freedom showed no difference, $\chi^2(2) = 4.78, p = .0919$. Therefore, we adhere to the more parsimonious Model A1.

Model B: Including persistence and the interaction between persistence and time

Model B tested whether the chess ratings of the persisters and drop-outs developed differently over time. Therefore, the factor persistence and the interaction between persistence and time were entered in the formula. The results showed that the coefficient for the factor persistence was not a significant contributor to the equation, but that the interaction between persistence and time was. The positive regression weight for this interaction indicates that persisters' chess ratings improved more than drop-outs' ratings. The AIC and BIC of this model are considerably lower than of Model A1: 19.63 and 9.9, respectively. This indicates an improvement of Model B over Model A1. This conclusion was further supported by a likelihood ratio test, comparing Model A1 and B, which resulted in a significant change to the model, $\chi^2(2) = 23.62, p < .0001$. Therefore, we concluded that both the factor persistence and the interaction between persistence and time led to a more informative model, and should be included in further analyses. Note that the persistence coefficient in Model B1 had a negative value, which indicates that persisters had overall lower chess ratings than drop-outs. One should, however, take into account that the drop-outs were on average 2.52 years older and had therefore practiced more, which explains their higher chess ratings. This is substantiated by the fact that, when the interaction between persistence and time is taken into the analysis, this negative effect disappears, and a positive relation between persistence and chess ratings is found.

To study whether persisters and drop-outs differentially benefited from deliberate practice activities, we entered the interaction between persistence and serious chess play in Model B2, and the interaction between persistence and serious chess study alone in Model B3. For Model B2 (serious chess play), this regression weight was significant. For Model B3 (serious chess study alone), the regression weight was not significant. To compare the models B1 through B3, we first examined the absolute value of the AIC and BIC. This comparison showed that the model that included the interaction between persistence and time (Model B1) explained the data best. Furthermore, the Akaike weights and Schwartz weights revealed that Model B1 had by far the highest probability among the three tested models (Akaike weight B1 = .9924, B2 = .0075, B3 = .00001; Schwartz weight B1 = .9924, B2 = .0075, B3 = .00001). Although the significant, negative interaction between persistence and serious play in Model B2 indicates that persisters seemed to profit less from serious play than drop-outs, the AIC and BIC values led us to conclude that Model B1 was the most appropriate explanation for the data. Apparently, hours of serious chess study alone and serious chess play did not differentially affect the chess ratings of persisters and drop-outs. Stated otherwise, both persisters and drop-outs benefited to the same extent from serious chess study alone and serious chess play (i.e. deliberate practice). Therefore, we decided to exclude these interactions from further analyses and only incorporate the factor persistence and the interaction between persistence and time (Table 5).

Model C: Effect of gender

The last model examined the effect of gender, and the interaction between gender and time, after the deliberate practice variables, time, and persistence were entered in the analysis. Again, the AIC (11,167.02), BIC (11,215.69), and a likelihood ratio test proved that adding gender and the interaction between gender and time led to an improvement of the model, $\chi^2(2) = 24.40$, $p < .0001$ (Table 6).

We also examined whether males and females profited differently from deliberate practice, by adding the interaction between gender and either serious chess play or serious chess study in Model C2 and C3, respectively. The AIC indicated that the model with the interaction between gender and time explained the data best. The Akaike weights and Schwartz weights led to the same conclusion (Akaike weights C1 = .8873, C2 = .0518, C3 = .0608; Schwartz weights C1 = .8873, C2 = .0518, C3 = .0608). Therefore, the interaction between gender and serious chess play and gender and serious chess study alone were not incorporated in the final model. The final model was of the form:

$$y_{ij} = \beta_0 + v_{0i} + \beta_1 T_{1ij} + \beta_2 P_{2ij} + \beta_3 S_{3ij} + \beta_4 D_{4ij} + \beta_5 D_{4ij} T_{1ij} + \beta_8 G_{5ij} + \beta_9 G_{5ij} T_{1ij}.$$

This model included the factor time, serious chess play, serious chess study alone, persistence, the interaction between persistence and time, gender, and the interaction between gender and time to account for variance in chess ratings.

To further study the relation between persistence and deliberate practice, and to provide more information on the effect the age difference between persisters and

Table 5. Estimated coefficients of model B (and standard errors), which tests the influence of serious chess play, chess study, persistence, and interactions between these variables, on chess performance over time

	Model B1		Model B2		Model B3	
Coefficient	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
Fixed parameters						
β_0 (constant)	1,591.98**	31.63	1,547.84**	32.53	1,566.13**	32.20
β_1 (T_1, time)	6.32**	0.16	6.83**	0.12	6.79**	0.12
β_2 (P_2, serious play)	0.05*	0.02	0.13**	0.03	0.06**	0.02
β_3 (S_3, serious study)	0.11**	0.02	0.13**	0.02	0.15**	0.04
β_4 (D_4, persistence)	− 17.83	40.00	43.12	41.85	13.40	41.00
$\beta_5(D_4 \times T_1)$	0.91**	0.23				
$\beta_6(D_4 \times P_2)$			−0.11**	0.036		
$\beta_7(D_4 \times S_3)$					−0.036	0.04
R	0.710		0.702		0.705	
Random parameters						
σ^2_{vi0}	169.28	65.49	170.86	65.67	170.01	66.01
Number of parameters	7		7		7	
AIC	11,187.42		11,197.21		11,205.61	
BIC	11,226.37		11,236.16		11,244.56	
Log likelihood	−5,585.71		−5,590.61		−5,594.80	

Note. *T* is the number of months from study entry at which the measurements were taken. P_2 (serious play) and S_3 (serious study) were estimated per year. $D_4 = 0$ for those who eventually dropped out, 1 for those who were still in national training at the time of study. Number of participants = 78, number of measurements = 968. *$p < .05$; **$p < .01$. BIC was based on the total number of measurements over all participants (968).

drop-outs might have played, we ran two extra models in which we predicted either the accumulated amount of serious chess play (Model D1) or the accumulated amount of serious chess study (Model D2) by the variables persistence, time, age, and the interaction between persistence and time. These models only differ in the dependent variable that was predicted. Both models showed the same results (See Table 7): only time and the interaction between persistence and time contributed to the equation. This indicates that the total amount of deliberate practice was not influenced by age or persistence itself, but was a result of differences in practice hours between persisters and drop-outs that appeared and increased as time passed. This is also illustrated by the graphical representation of these data in Figure 1. This Figure represents the mean hours of serious chess play and serious chess study alone for the first 10 years of seriously playing chess. The graph shows a marked increase in practice hours across time, and a difference between persisters and drop-outs that increases over time.

Finally, all tested models showed higher coefficients for serious chess study than for serious chess play. We examined whether this difference was significant and 1 hour of serious chess study indeed had a stronger effect on chess rating than 1 hour of serious play. Since in linear mixed model analysis no methods exist to compare beta weights, we used the Hotelling–Williams statistic (Williams, 1959) that is applied in traditional regression analyses. In a model with chess rating as dependent variable, and serious chess play and serious chess study as independent variables, a Hotelling–Williams

Table 6. Estimated coefficients of model C (and standard errors), which tests the influence of serious chess play, chess study, persistence, gender, and interactions between these variables, on chess performance over time

	Model C1		Model C2		Model C3	
Coefficient	Estimate	*SE*	Estimate	*SE*	Estimate	*SE*
Fixed parameters						
β_0 (constant)	1,518.83**	38.64	1,485.97**	40.89	1,495.39**	39.11
β_1 (T_1, time)	5.86**	0.26	6.33**	0.16	6.32**	0.16
β_2 (P_2, serious play)	0.05*	0.02	0.12*	0.05	0.05*	0.02
β_3 (S_3, serious study)	0.10**	0.02	0.10**	0.02	0.20**	0.06
β_4 (D_4, persistence)	− 16.14	37.54	− 16.23	37.41	− 16.63	41.00
$\beta_5(D_4 \times T_1)$	0.93**	0.23	0.90**	0.23	0.92**	0.23
β_8 (G_5, gender)	118.09**	38.40	156.48**	41.34	147.01**	39.22
$\beta_9(G_5 \times T_1)$	0.59*	0.27				
$\beta_{10}(G_5 \times P_2)$			− 0.09	0.06		
$\beta_{11}(G_5 \times S_3)$					− 0.10	0.06
R	0.749		0.751		0.751	
Random parameters						
σ^2_{vi0}	158.39	65.34	157.82	65.45	157.77	65.45
Number of parameters	9		9		9	
AIC	11,167.02		11,172.70		11,172.38	
BIC	11,215.69		11,221.37		11,221.05	
Log likelihood	− 5,573.51		− 5,576.36		− 5,576.12	

Note. T is the number of months from study entry at which the measurements were taken. P_2 (serious play) and S_3 (serious study) were estimated per year. $D_4 = 0$ for those who eventually dropped out, 1 for those who were still in national training at the time of study. $G_5 = 0$ for females, 1 for males. Number of participants = 78, number of measurements = 968. $*p < .05$; $**p < .01$. BIC was based on the total number of measurements over all participants (968).

statistic of 0.99 was achieved ($p = .32$), which is not significant at the .05 level. Therefore, we conclude that the difference in beta weights of serious chess study and serious chess play is not significant.

Discussion

This study set out to explore the development of the relation between deliberate practice and performance in chess over time. More specifically, we examined (a) to what extent the monotonic benefits assumption holds when tested over time, (b) whether persisters and drop-outs benefit to the same extent from deliberate practice, and (c) the influence of deliberate practice on performance differences between males and females. We found that serious chess study alone and serious chess play against others were strong contributors to performance, irrespective of the moment in chess players' careers. Moreover, our results revealed that differences in performance between persisters and drop-outs were not caused by differential benefit from deliberate practice. That is, adding the interaction between persistence and either serious chess study or serious chess play led to a less informative model. The same line of reasoning applied to male–female performance differences. These could mainly be explained by

Table 7. Estimated coefficients of model D (and standard errors), which tests the effect of persistence, age and time on accumulated hours of serious chess play (Model D1), or accumulated hours of serious chess study (Model D2)

	Model D1		Model D2	
Coefficient	Estimate	*SE*	Estimate	*SE*
Fixed parameters				
β_0 (constant)	−320.45	543.52	−30.25	631.73
β_1 (T_1, time)	27.21**	0.98	25.61**	1.17
β_2 (D_4, persistence)	−9.59	165.44	−14.51	192.45
$\beta_3(D_4 \times T_1)$	9.55**	1.33	12.21**	1.58
$\beta_4(A_5$, Age)	16.66	29.60	−10.49	34.40
R	0.676		0.664	
Random parameters				
σ^2_{vi0}	593.35	397.31	688.57	471.32
Number of parameters	6		6	
AIC	14,560.49		14,886.37	
BIC	14,594.58		14,920.46	
Log likelihood	−7,273.25		−7,436.18	

Note. *T* is the number of months from study entry at which the measurements were taken. $D_4 = 0$ for those who eventually dropped out, 1 for those who were still in national training at the time of study. A_5 stands for age. Number of participants = 78, number of measurements = 968. $^{**}p < .01$. BIC was based on the total number of measurements over all participants (968).

the lower amount of deliberate practice by females and not by a smaller effect of deliberate practice on performance.

The current study provides support for the monotonic benefits assumption, which states that current performance is monotonically related to accumulated amount of deliberate practice. Moreover, the findings indicate that, in chess, this assumption not only applies to current performance, but also to the relation between deliberate practice and performance over time. That is, regardless of age and performance level of chess players, deliberate practice is monotonically associated with performance throughout chess career. Our findings suggest that, in chess, deliberate practice has an immediate effect on performance, at least when measured at 1-year intervals. Although peak performance in chess usually takes place in the mid-thirties or even later (Charness & Bosman, 1990; Roring & Charness, 2007), our study shows that from around 12 years on, chess players' engagement in deliberate practice increases steadily, and, concurrently, chess ratings improve. These findings emphasize the dedicated and structured manner in which young chess players approach their favourite pastime.

As to the question at what stage in career deliberate practice is determining of performance, our results suggest that, in chess, deliberate practice plays a crucial role very early on, and to the same extent throughout career. These findings diverge from observations in studies in athletics, which showed that sport-specific practice was not important until later in athletes' careers (e.g. Baker *et al.*, 2003). However, a difference in setting might explain this discrepancy in results. In athletics, individuals usually start by engaging in a wide range of sport activities (i.e. the 'sampling years', Côté, Baker, & Abernethy, 2001), before becoming devoted to a single sport (i.e. the 'specializing years'). In the sampling years, general motor skills are developed that will prove

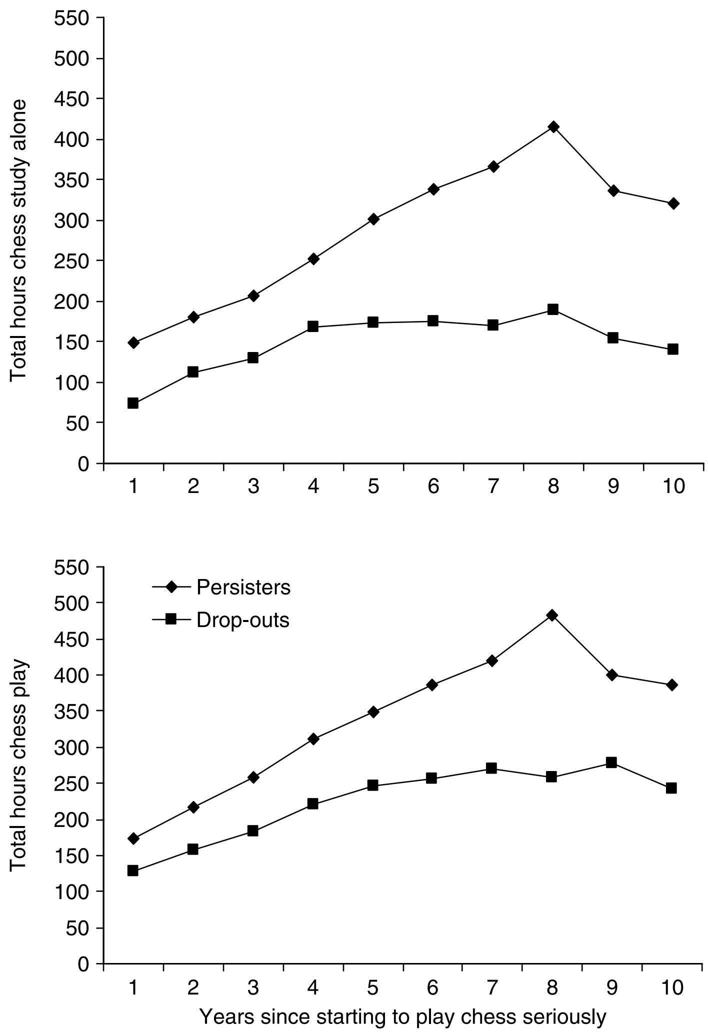

Figure 1. Number of hours of serious chess study alone and serious chess play per year separated for persisters and drop-outs, beginning at the year that they seriously started playing chess.

beneficial to performance in the investment years. Therefore, although practice in the first years is not specific to the sport in which the individual will become an expert, it does have an effect on later performance. As a result, early sport-specific practice is not a necessary condition for expertise, and can be substituted by practice in related sports. Because there is no evidence that general cognitive skills need to be developed before learning to play chess, domain-specific practice plays a more crucial role and does not overlap with other activities. Therefore, early chess-specific practice shows a more direct link to chess performance than early sport-specific practice to performance in athletics.

Our results revealed that those who dropped out of the national training did not differ *a priori* from the persisters in chess ratings, but in development of chess ratings over time. That is, both groups had similar chess ratings at the time of selection for the national training, but as time passed, the ratings of the drop-outs began to lag behind ratings of the persisters. However, despite the slower growth in chess ratings of the drop-outs, persisters and drop-outs benefited to the same extent from deliberate practice, as indicated by the model that best explained the data and that did not include the interactions between persistence and deliberate practice. This suggests that the lower chess ratings of the drop-outs are not caused by profiting less from investments in deliberate practice than the persisters, but by spending less time on deliberate practice. This is the first study that has tested the critique raised against the deliberate practice theory that the factors talent and deliberate practice could be confounded (Sternberg, 1996). Our results indicate that those who ultimately arrive at expert level in chess do so not because of a predisposition to perform deliberate practice more efficiently, but because they put in more hours of deliberate practice. Although all participants were selected for the Dutch national chess training, the variance in chess ratings in this group was considerable. Moreover, our sample was young and relatively inexperienced compared to other samples in comparable research (e.g. Charness *et al.*, 1996, 2005). This suggests that even within skill levels and regardless of absolute performance, differences in performance could mainly be attributed to variation in time dedicated to deliberate practice throughout career. This finding provides implications for training: irrespective of skill level, stimulating deliberate practice will likely improve performance. Further research should focus at testing this assumption in other domains besides chess. Moreover, research should study chess players who only recently started playing chess and follow their practice behaviour and performance development to provide a prospective examination of this relation.

As to the influence of gender on chess performance, our results provide more support for an effect of practice differences between males and females than for the participation rate hypothesis (Chabris & Glickman, 2006). According to Charness and Gerschak's (1996) formula, the difference between males and females in our sample would have been 39 rating points based on the proportion of females in the sample, whereas the mean difference was in fact 203 rating points. However, this performance difference could largely be explained by variation in deliberate practice. The non-significance of the interactions between gender and serious chess play, and gender and serious chess study alone indicates that both genders profit equally from deliberate practice. Apparently, similar investments in deliberate practice lead to similar performance improvements for both males and females. However, the results also reveal that, even after controlling for differences in amount of deliberate practice, males tend to have a slight advantage in chess ratings over females. As our findings indicate that males and females differ to a large extent in hours dedicated to deliberate practice, and in variables as number of chess books and CDs owned, we cannot rule out the possibility that other differences in practice-related variables exist that are not measured here, but that can further explain the performance difference between males and females. A recent study by Maass, D'Ettole, and Cadinu (in press), for example, provides support for the negative effect of stereotype threat on female performance in chess. Further research is needed that more minutely analyses the differences in chess activities between males and females. In general, our results provide a more moderate view on the effect of gender on performance compared to Howard's (2004) conclusion. We show that, when explaining variation in chess performance, variation in deliberate practice is a strong determinant.

The current findings provide possible implications for training and education of chess players. Since our study did not aim to explore motivational differences between males and females, or persisters and drop-outs, we can not draw any conclusions about *why* females and drop-outs practiced less and, in the latter case, eventually quit. Our data do indicate that both females and drop-outs were not characterized by profiting less from practice, or do our data allow for any alternative explanations for performance differences that rely on unchangeable biological causes. However, our findings mainly provide support for the strong influence of deliberate practice on chess performance. Therefore, we would recommend, particularly for females and for those who are thinking of quitting, an intervention aimed at increasing practice intensity. This would have the needed effect on performance, and thereby increase motivation to practice more and continue playing chess.

In sum, the present study provides new evidence for the viability of the deliberate practice theory in chess. For the first time, the validity of the monotonic benefits assumption was demonstrated longitudinally, by revealing that groups of different performance development profit similarly from deliberate practice. In future investigations, this assumption needs to be examined in non-cognitive domains. As these results imply that deliberate practice also induces performance improvement in non-experts, further research should focus attention on teaching the characteristics of deliberate practice to individuals of lower expertise. Finally, we detected that gender has a relatively small effect on performance above the effect of deliberate practice. We encourage use of the mixed models analysis in deliberate practice research in other fields besides chess, in order to extend and refine our knowledge of the effect of deliberate practice on the development of expertise.

Acknowledgements

The authors would like to thank Jeroen Bosch and Esther de Kleuver for their help in designing the questionnaire and recruiting participants, and Mirjam Laging for her help in collecting the data. Furthermore, the authors would like to thank Gino Camp and two anonymous reviewers for helpful comments on an earlier draft of this manuscript.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Caski (Eds.), *Proceedings of the second international symposium on information theory* (pp. 267–281). Budapest: Akademiai Kiado.

Baker, J., Côté, J., & Abernethy, B. (2003). Sport specific training, deliberate practice, and the development of expertise in team ball sports. *Journal of Applied Sport Psychology*, *15*, 12–25.

Baker, J., Côté, J., & Deakin, J. (2005). Expertise in ultra-endurance triathletes early sport involvement, training structure, and the theory of deliberate practice. *Journal of Applied Sport Psychology*, *17*, 64–78.

Benbow, C. P., Lubinski, D., Shea, D. L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability at age 13: Their status 20 years later. *Psychological Science*, *11*, 474–480.

Bilalić, M., & McLeod, P. (2006). How intellectual is chess? A reply to Howard. *Journal of Biosocial Science*, *39*, 1–3.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical-theoretic approach* (2nd ed.). New York: Springer.

Chabris, C. F., & Glickman, M. E. (2006). Sex differences in intellectual performance. *Psychological Science*, *17*(12), 1040–1046.

Charness, N., & Bosman, E. A. (1990). Expertise and aging: Life in the lab. In T. H. Hess (Ed.), *Aging and cognition: Knowledge organization and utilization* (pp. 343–385). Amsterdam: Elsevier.

Charness, N., & Gerschak, Y. (1996). Participation rates and maximal performance: A log-linear explanation for group differences, such as Russian and male dominance in chess. *Psychological Science*, *7*, 46–51.

Charness, N., Krampe, R., & Mayr, U. (1996). The role of practice and coaching in entrepreneurial skill domains: An international comparison of life-span chess skill acquisition. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games* (pp. 51–80). Mahwah, NJ: Erlbaum.

Charness, N., Tuffiash, M., Krampe, R., Reingold, E., & Vasyukova, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology*, *19*(2), 151–165.

Côté, J., Baker, J., & Abernethy, B. (2001). Stages of sport participation of expert decision-makers in team ball sports. In A. Papaionnou, M. Goudas, & Y. Theodorakis (Eds.), *In the dawn of the new millennium – program and proceedings of the 10th world congress of sport psychology* (Vol. 3, pp. 150–152). Thessaloniki, Hellas: Christodoulidi Publishing.

De Bruin, A. B. H., Rikers, R. M. J. P., & Schmidt, H. G. (2007). The influence of achievement motivation and chess-specific motivation on deliberate practice. *Journal of Sport and Exercise Psychology*, *29*(5), 561–583.

Doll, J., & Mayr, U. (1987). Inelligenz und Schachleistung – eine Untersuchung an Schachexperten {Intelligencen and performance in chess – A study of chess experts}. *Psychologische Beiträge*, *29*, 270–289.

Duffy, L. J., Baluch, B., & Ericsson, K. A. (2004). Dart performance as a function of facets of practice amongst professional and amateur men and women players. *International Journal of Sport Psychology*, *35*, 232–245.

Dunn, T. G., & Shriner, C. (1999). Deliberate practice in teaching: What teachers do for self-improvement. *Teaching and Teacher Education*, *15*, 631–651.

Ericsson, K. A. (1996). The acquisition of expert performance: An introduction to some of the issues. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games* (pp. 1–50). Mahwah, NJ: Erlbaum.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, *49*(8), 725–747.

Ericsson, K. A., Krampe, R. Th., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, *47*, 273–305.

Frydman, M., & Lynn, R. (1992). The general intelligence and spatial abilities of gifted young Belgian chess players. *British Journal of Psychology*, *83*, 233–235.

Geiselman, R. E., Fisher, R. P., Firstenberg, I., Hutton, L. A., Sullivan, S., Avitissian, I., *et al.* (1984). Enhancement of eyewitness memory: An empirical evaluation of the cognitive interview. *Journal of Police Science and Administration*, *12*, 74–80.

Gustin, W. C. (1985). The development of exceptional research mathematicians. In B. S. Bloom (Ed.), *Developing talent in young people* (pp. 270–331). New York: Ballantine Books.

Helsen, W. F., Starkes, J. L., & Hodges, N. J. (1998). Team sports and the theory of deliberate practice. *Journal of Sport and Exercise Psychology*, *20*, 12–34.

Hodge, T., & Deakin, J. M. (1998). Deliberate practice and expertise in the martial arts: The role of context in motor recall. *Journal of Sport and Exercise Psychology*, *20*, 260–279.

Hodges, N. J., Kerr, T., Starkes, J. L., Weir, P. L., & Nananidou, A. (2004). Predicting performance times from deliberate practice hours for triathletes and swimmers: What, when, and where is practice important? *Journal of Experimental Psychology: Applied*, *10*(4), 219–237.

Hodges, N. J., & Starkes, J. L. (1996). Wrestling with the nature of expertise: A sport-specific test of Ericsson, Krampe, and Tesch-Römer's (1993) theory of 'deliberate practice'. *International Journal of Sport Psychology*, *27*, 400–424.

Howard, R. (2004). Are gender differences in high achievement disappearing? A test in one intellectual domain. *Journal of Biosocial Studies*, *37*, 371–380.

Krampe, R. T., & Ericsson, K. A. (1996). Maintaining excellence: Deliberate practice and elite performance in younger and older pianists. *Journal of Experimental Psychology: General*, *125*(4), 331–359.

Laird, N. M., & Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics*, *38*, 963–974.

Maass, A., D'Ettole, C., & Cadinu, M. (2008). Checkmate? The role of gender stereotypes in the ultimate intellectual sport. *European Journal of Social Psychology*, *38*(2), 231–245.

Memon, A., & Bull, R. (1991). The cognitive interview: Its origins, empirical support, evaluation, and practical implications. *Journal of Community and Applied Social Psychology*, *1*, 291–307.

Monsaas, J. A. (1985). Learning to be a world class tennis player. In B. S. Bloom (Ed.), *Developing talent in young people* (pp. 211–269). New York: Ballantine Books.

Omar, R. Z., Wright, E. M., Turner, R. M., & Thompson, S. G. (1999). Analysing repeated measurements data: A practical comparison of methods. *Statistics in Medicine*, *18*, 1587–1603.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-Plus*. New York: Springer.

Roring, R. W., & Charness, N. (2007). A multilevel model analysis of expertise in chess across the life span. *Psychology and Aging*, *22*(2), 291–299.

Shie, Y. Y., & Fouladi, R. T. (2003). The effect of multicollinearity on multilevel modelling parameter estimates and standard errors. *Educational and Psychological Measurement*, *63*, 951–985.

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist*, *61*, 394–403.

Sloboda, J. A., Davidson, J. W., Howe, M. J. A., & Moore, D. G. (1996). The role of deliberate practice in the development of performing musicians. *British Journal of Psychology*, *87*, 287–309.

Starkes, J. L., Deakin, J. M., Allard, F., Hodges, N. J., & Hayes, A. (1996). Deliberate practice in sports: What is it anyway? In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games* (pp. 81–106). Mahwah, NJ: Erlbaum.

Sternberg, R. J. (1996). Costs of expertise. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games* (pp. 347–354). Mahwah, NJ: Erlbaum.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin and Review*, *11*(1), 192–196.

Ward, P., Hodges, N. J., Williams, M. A., & Starkes, J. L. (2004). Deliberate practice and expert performance: Defining the path to excellence. In A. M. Williams & N. J. Hodges (Eds.), *Skill acquisition in sport: Research, theory, and practice*. London: Routledge.

Waters, A. J., Gobet, F., & Leyden, G. (2002). Visuospatial abilities of chess players. *British Journal of Psychology*, *93*, 557–565.

Williams, E. (1959). *Regression analysis*. New York: Wiley.

Received 9 August 2006; revised version received 22 February 2008